



Southeast Alaska Network

SEAN Freshwater Water Quality Cast Data User Guide

Abstract

Basic freshwater quality observations are maintained by SEAN in a consistent database. This paper summarizes the scope of these data. Methods for accessing data are explained. References are given for finding definitions of the collected attributes.

Table of Contents

1	The Origin of These Data	3
2	Availability of Detailed Data	3
3	Known Data Limitations	3
4	Mechanics of Accessing Data	3
5	Spatial Coverage.....	4
6	Temporal Coverage	5
7	Authority	5
8	Analysis Facilities	5
9	Further Information	5
10	Description of Data Fields.....	6

Revision history

Revision Date	Base Version	Author	Nature of Change
1/4/2012	-new-	Bill Johnson	Initial version

1 The Origin of These Data

Oceanographic data have been collected in the Glacier Bay area since 1993. Data points are obtained by lowering a CTD device through the water column and recording sensor readings at specific intervals. The set of observations collected at a particular station and time is referred to as a cast. Cast data are further stratified by depth. A sonde is fitted with detectors that measure such environmental factors as temperature, conductivity, turbidity, dissolved oxygen, and pH. Not all factors have been recorded at all times in all locations. For example, turbidity was only measured on the Taiya River.

The output of a sonde is captured as “raw data” in a proprietary format. To be useful, the raw data are saved in ASCII files. Each line of an ASCII file is augmented with metadata such as record quality, data grade, protocol taken under, etc. Much of the augmentation is taken from Site Visit Worksheets that record field conditions. Data and versioning metadata are then inserted into the cumulative database so interested parties may access them online.

2 Availability of Detailed Data

Data collection was begun in calendar year 2010 under a draft protocol, eventually named FQ-2011.1. The first year measured only Indian River, Sitka and Salmon River, Gustavus. In 2011 the Taiya River of Skagway was added. Specific data availability by parameter, location, and month are maintained in a [table on the SEAN web site](#).

Where particular data points were not collected (as, say, when a specific detector was not in service) those values are left null in the database. A single data structure holds all sonde data, and it may be queried in a consistent manner regardless of particular season, equipment used, or project staffing during data collection.

SEAN’s data are in the public domain.

3 Known Data Limitations

While not specifically a limitation, users must know that parameter values in the database come directly from the instruments and are not corrected for sensor drift or fouling. Since there are numerous methods for correcting data, no one of which is likely to best meet every purpose of every researcher, correcting for outside factors is left to the user. Note the supplied data grade values for each reading may be useful in determining which, if any, data are likely candidates for correction.

4 Mechanics of Accessing Data

Authoritative data are available through the SEAN web site at http://science.nature.nps.gov/im/units/sean/FQ_F_Browse.aspx. The purpose of this web page is to

provide access to the basic data. Results are available as both screen grids and file downloads. There is a button on the web page to generate each. Grids are populated with results and may be copied and pasted into other applications using standard Windows facilities. The grid is generally limited to 300 rows. This limit protects users from overloading their computers; there are potentially many rows available, each with many columns. In actual operation, it is usually convenient to adjust the filtering parameters until the grid contents confirm that desired data are being delivered. Then data sets of any size may be saved on one's workstation in comma separated value files by pressing the "Save Results to a CSV File" button.

Filtering may be done by specifying values in the appropriate fields on the page. Filter parameters restrict the result set. If filter parameters are omitted, then all values are included for that attribute.

A particular year may be selected, or all years. If a set of, say, three or four specific years are desired, then a separate download must be made for each. Alternatively, all years may be selected which is then narrowed down by specifying a range of dates that cross a span of multiple years.

A sampling date range may be specified. If earliest date is omitted, the first date ever sampled is used. If latest date is omitted, then today's date is used. Be aware that specifying a date range lying outside the chosen year will result in finding zero rows.

If no site is specified, all are returned. Check boxes may be used to restrict results to certain sites.

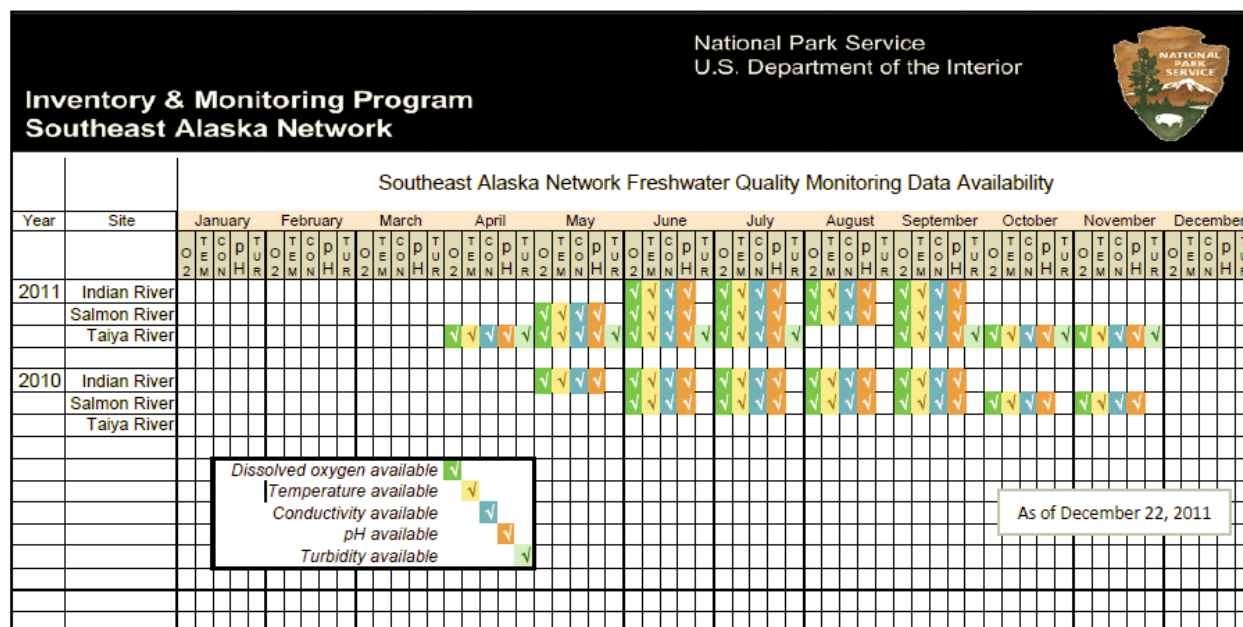
A check box is available for specifying whether or not to include comment fields in the extract. Because comments on each row can be lengthy it is sometimes more convenient to retrieve data without them.

5 Spatial Coverage

The current locations of observed sites may be found at http://science.nature.nps.gov/im/units/sean/FQ_Site_Map.aspx.

It is possible that new sites may be added to the program and old ones dropped over the years. In such cases a previously used site identifier is never reused. Original data are retained and available under the proper site identifier, regardless of whether or not a particular location is still being actively surveyed.

6 Temporal Coverage



Above is a snapshot of data coverage as of this writing. The most up-to-date version of this table is available on the web site under “Data Availability.”

7 Authority

The data browser pulls rows directly from SEAN’s vital sign database. These data reflect the authoritative source. Final analyses should always be based on data pulled from the browser. Other sources, such as spreadsheet files, paper tables typed back into computers, or email communications cannot be relied on to be authoritative.

SEAN is in the process of arranging for external repositories to house these data. Such repositories include the NPStoret program and IRMA portal. It should be noted that the maintenance of data copied to these repositories is out of the control of SEAN. If any data discrepancy is seen among the sources, the SEAN website should be considered the authority.

8 Analysis Facilities

The data browsing feature gives everyone the ability to extract data of interest onto their local computers. From there they are free to use whatever tools they choose to perform whatever analyses they are interested in. The data browser does not provide embedded analytical tools.

9 Further Information

Additional information, as well as current contact persons, may be found in the freshwater water quality section of <http://science.nature.nps.gov/im/units/sean/>.

10 Description of Data Fields

Formal metadata, including brief field descriptions, are available on the SEAN web site at http://science.nature.nps.gov/im/units/sean/FQ_F_Browse.aspx . More detailed explanations are provided in the protocol document itself, available on SEAN's main freshwater water quality web page. Note that each row in the database is stamped with the protocol it was collected and validated under, and different protocols may treat specific fields somewhat differently. Be sure to select the proper protocol document when investigating the fine details of data fields.